DOCUMENT RESUME

ED 210 696 CS 206 673

AUTHOR Mellon, John
TITLE Sentence-Combining Skills: Results of the Sentence-Combining Exercises in the 1978-79 National Writing Assessment.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY National Center for Education Statistics (DHEW), Washington, D.C.; National Inst. of Education (ED), Washington, D.C.
REPORT NO NAEP-R-10-W-65
PUB DATE Dec 81
CONTRACT OEC-0-74-0506
GRANT NIE-G-80-0003
NOTE 64p.; Table A-1 may not be legible due to small print.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; Evaluation Methods; National Surveys; *Sentence Combining; *Writing Evaluation; *Writing Instruction; *Writing Research
IDENTIFIERS *National Assessment of Educational Progress; Third Writing Assessment (1979)

ABSTRACT

This report presents the results of a battery of sentence combining tasks included in the assessment of writing conducted by the National Assessment of Educational Progress in 1978-79. The first section of the report provides background information about the NAEP survey and discusses the procedures used to develop and score the sentence combining tasks. The second section presents the 14 sentence combining tasks that were administered to 9-, 13-, and 17-year-old students across the United States, along with the most acceptable combination, and the percentage of students who produced acceptable responses. The third section offers a breakdown of the results according to geographical region, sex, race, education level of parents, and type of community. It also relates student performance on the sentence combining tasks to their primary trait or holistic essay scores and to syntactic descriptions of the sentences comprising their essays. The fourth section provides a summary of the results. Appendixes contain frequency rankings of correct papers and descriptions of the writing exercises and scoring procedures. Primary type of information provided by report: Results (Secondary Analyses) (Interpretation). (FL)

SENTENCE-COMBINING SKILLS

Results of the Sentence-Combining Exercises in the 1978-79 National Writing Assessment

Special Paper No. 10-W-65

Prepared by
John Mellon

for the
National Assessment of Educational Progress

Education Commission of the States
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295

December 1981

The National Assessment of Educational Progress is funded by the National Institute of Education. It is under contract with the Education Commission of the States. It is the policy of the Education Commission of the States to take affirmative action to prevent discrimination in its policies, programs and employment practices.

The National Assessment of Educational Progress is an education research project mandated by Congress to collect and report data, over time, on the performance of young Americans in various learning areas. National Assessment makes available information on assessment procedures and materials to state and local education agencies and others.

The work upon which this publication is based was performed pursuant to Contract No. OEC-0-74-0506 of the National Center for Education Statistics and the National Institute of Education; also, Grant No. NIE-G-80-0003 of the National Institute of Education. It does not, however, necessarily reflect the views of those agencies.

TABLE OF CONTENTS

LIST OF TABLES

FOREWORD

When the U.S. Office of Education was chartered in 1867, one charge to its commissioners was to determine the nation's progress in education. The National Assessment of Educational Progress (NAEP) was initiated a century later to address, in a systematic way, that charge.

Since 1969, the National Assessment has gathered information about levels of educational achievement across the country and reported its findings to the nation. It has surveyed the attainments of 9-year-olds, 13-year-olds, 17-year-olds and adults in art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies and writing. All areas have been periodically reassessed in order to detect any important changes. To date, National Assessment has interviewed and tested nearly 1,000,000 young Americans.

Learning-area assessments evolve from a consensus process. Each assessment is the product of several years of work by a great many educators, scholars and lay persons from all over the nation. Initially, these people design objectives for each subject area, proposing general goals they feel Americans should be achieving in the course of their education. After careful review, these objectives are given to writers, whose task it is to create exercises (items) appropriate to the objectives.

When the exercises have passed extensive reviews by subject-area specialists, measurement experts and lay persons, they are administered to probability samples. The people in these samples are selected in such a way that the results of their assessment can be generalized to an entire national population. That is, on the basis of the performance of about 2,500 9-year-olds on a given exercise, we can make generalizations about the probable performance of all 9-year-olds in the nation.

After assessment data have been collected, scored and analyzed, the National Assessment publishes reports and disseminates the results as widely as possible. Not all exercises are released for publication. Because NAEP will readminister some of the same exercises in the future to determine whether the performance levels of Americans have increased, remained stable or decreased, it is essential that they not be released in order to preserve the integrity of the study.

CHAPTER 1

BACKGROUND

A. Overview

This report presents the results of a battery of sentence-combining (SC) tasks included for the first time by the National Assessment of Educational Progress (NAEP) in its third national assessment of writing, conducted in 1978-79. Students aged 9, 13 and 17 performed these SC tasks in addition to writing essays in response to a variety of topics and to answering questions about their writing experiences and attitudes. NAEP Reports 10-W-01, 10-W-02 and 10-W-03, Writing Achievement, 1969-79: Results From the Third National Writing Assessment, present in three volumes the results of the writing exercises and the experience and attitude questionnaires for all age groups in the three assessments thus far conducted.

This report analyzes performance on the SC tasks and relates that performance to the students' primary-trait or holistic essay scores, and to syntactic descriptions of the sentences comprising their essays. Coming as it does at the end of a decade that saw SC practice become fairly widespread in school and college classrooms, and that produced a number of research studies verifying the effectiveness of SC in enhancing the acquisition of syntactic skills, National Assessment's incorporation of SC tasks into a nationwide assessment of writing represents an extension of the domain of SC methodology that is both timely and significant.

Why Assess SC Ability?

Described as nontechnically as possible, sentence combining refers to the mostly automatic and unconscious use of grammatical operations (transformations) that enable a writer to include in each full sentence all the ideas (i.e., all the elementary propositions of thought, often called "kernel sentences") the writer needs to include in order to say what he or she wishes to say. Because most sentences necessarily consist of kernels that writers have combined in gramatically acceptable ways, SC principles come into play in everyone's language productions, whether or not they have had formal grammar study or practiced sentence combining in school. The purpose of a sentence-combining practice in school is to make students more conscious of the

choices available for expressing their ideas than they otherwise might be, and to give them experience in exercising these choices in sentences more mature in structure than they otherwise might write.

The advent of transformational grammar in the 1960s gave writing researchers a new and illuminating theoretical frame of reference through which to study the development of syntactic maturity in the writing of school-age youth. At the same time, pedagogical researchers began experimenting with various forms of SC exercises in grammar and writing curricula. Sentence-combining skill, whether acquired exclusively in naturalistic ways, through practice in language use or also specifically taught, came to be termed "syntactic fluency." Syntactic fluency is now generally regarded as a separately definable and measurable aspect of writing ability.

The Assessment's decision to measure sentence combining by direct means follows from its commitment to describe relationships between global characteristics of pieces of writing, as measured by primary rhetorical trait (PT) and holistic (overall quality) scoring, and componential factors of the whole, such as mechanical conventions, devices of cohesion, types of idea groupings and organizing strategies, and, in the case of syntactic fluency, sentence-combining transformations.

Moreover, by using specifically designed SC tasks, instead of merely cataloguing SC transformations observed in free writing, National Assessment has sought to assess SC skill on a competence basis (what writers <u>can</u> do when specifically put to the task), as distinct from assessment on a performance basis (what writers <u>happen</u> to do when writing actual essays).

Within this context, there are several additional points readers of this report should bear in mind. First, because the present assessment represented the first attempt to use SC tasks in a writing assessment, the overall scope of sentence structures included was purposely kept quite modest and covers only some of the many transformations writers command. Second, it should be understood that the format of NAEP's SC exercises is not identical to that pioneered by any of the educational researchers whose SC materials have been made available through academic or commercial publications. Third, it is important to remember that NAEP's use of SC tasks as a type of assessment exercise was intended to be completely independent of questions as to whether certain students may have received instruction in pedagogical SC in their schools, and in no way constitutes an endorsement of any particular form of classroom sentence-combining practice.

B. Populations Assessed

The target population for each of the three assessments of writing consisted of 9-, 13- and 17-year-olds attending public or private schools.[1] Details of the sampling design and procedures are explained in NAEP Report 10-W-40, Procedural Handbook: 1978-79 Writing Assessment (1980), and numerous other Assessment reports and monographs. Here it should be sufficient to say that each assessment employed a stratified, multistage probability sample design. About 2,000-2,700 responses were collected for any given writing task. Some of the figures given in this report are based on an analysis of all 2,000-2,700 responses to a particular exercise, and some are based upon national subsamples of 365-722 papers -- a number sufficiently large to permit generalizations about an entire age group, but not large enough to permit statements about special subpopulations such as rural youngsters of a particular age. To obtain the representative subsamples of descriptive and narrative papers, scientific probability subsamples were drawn from the total National Assessment samples. Small percentages (1%-5%) of these subsamples were nonratable papers that were excluded from the analysis.

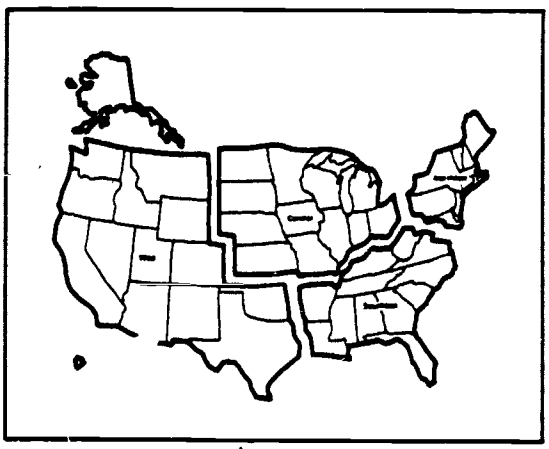
Whenever analysis is based upon full samples of 2,000-2,700, the National Assessment can report results for a number of population groups defined by sex, race, region of the country, parental education, type of community and grade in school. These are defined in Table 1-1.

---

[1]Although National Assessment included out-of-school 17-year-olds in the 1969 and 1974 assessments, resources did not permit data collection from this group in 1979. Since this report is concerned with changes over time, results are only presented for 17-year-olds attending school for each assessment.

TABLE 1-1. Definitions of Subgroups

Sex	Results are presented for males and females.
Race	Results are presented for black students and white students. Data for Hispanic students are not reported because sample sizes for individual items are too small.
Region	Results are presented for Northeastern, Southeastern, Central and Western regions shown on the following map.



Parental education	Results are presented for three levels of parental education: (1) those whose parents did not graduate from high school, (2) those who have at least one parent who graduated from high school and (3) those who have at least one parent who has had some post high school education.
Type of community	Three extreme community types of special interest are defined by an occupational profile of the area served by a school, as well as by the size of the community in which the school is located. This is the only reporting category that excludes a large number of respondents. About two-thirds do not fall into the classifications listed below. Results for the remaining two-thirds are not reported, since their performance is similar to that of the nation.

Advantaged urban. Students in this group attend schools in or around cities having a population greater than 200,000 and where a high proportion of the residents are in professional or managerial employment.

Disadvantaged urban. Students in this group attend schools in or around cities having a population greater than 200,000 and where a relatively high proportion of the residents are on welfare or not regularly employed.

Rural. Students in this group attend schools in areas with a population under 10,000 and where many of the residents are farmers or farm workers.

Grade in school

Results are presented for 17-year-olds in grades 10 (13%), 11 (75%) and 12 (11%). They are also available for 13-year-olds in grades 7 (25%), 8 (73%) and for 9-year-olds in grades 3 (25%) and 4 (72%).

C. SC Exercise Development and Scoring

In order to keep within the limits of available administration and scoring time, and also to make use of those transformational structures on which most pedagogical SC research has focused, NAEP exercise developers and professional consultants decided at the outset to confine the first SC assessment to single-sentence rather than whole-discourse exercises, and to the structures of noun modification and nominalization, to the exclusion of the structures of coordination, parallelism and logical conjunction. This meant that the exercises to be developed would aim to measure writers' ability to use various combinations of relative clauses and their pre- and postnoun reductions, along with nominal constructions such as factive clauses, indirect-question clauses, gerund phrases, and so on, although in fact all kinds of structures would be observed in the writers' responses.

Since the use of sentence-combining tasks as an assessment device represented a voyage into uncharted waters, arious exercise formats (usually referred to as "cues" or "signals") and sets of directions for working the exercises were developed and field tested. From the outset it became clear that only the simplest of formats and directions could be used. In the end, the following disarmingly simple language and unsignaled illustration exercise were used in all exercises:

Below are some sets of short sentences. Each set can be improved by combining the given sentences into one sentence that says the same thing. For example, if the sentences were:

A cat chased the ball.
The cat was big.
It was gray.

You could write:

The big gray cat chased the ball.

After you hear each set read aloud, read the sentences silently to yourself and figure out a way to combine them into one sentence. Be sure your sentence has the same meaning as the sentences in the given set. Then write your sentence on the lines.

One notes immediately that the task to be performed turns entirely on the phrase "can be improved." This has the effect of representing SC activity as an editing exercise, something a writer does to combine short and immature sentences occurring in his or her first-draft writing. In fact, however, as persons involved with pedagogical SC know very well, SC is never presented to students as an editing activity in normal classroom instruction; for if the minimal (kernel or basic) sentences found in SC exercises have any psychological reality at all, it is that they represent preconscious propositions of thought as they might occur in the mind prior to their first composition into ordinary surface-structure sentences -- that is, into the surface structures we "hear" in our heads while engaged in the act of writing. They are never presented as the sentences of first-draft writing, since not even 9-year-olds' prose has the artificial "baby talk" character of the kernel sentences used in SC exercises. Of course, no harm is done by presenting SC tasks used for assessment as if they were editing exercises, so long as persons interpreting the assessment bear in mind that pedagogical SC is not a matter of editing.

It should also be noted that the simplified directions finally used definitely limited what the exercises could cover. Nominalizations, for example, could not be specifically cued, and there was no way to designate which sentence should serve as the main clause in the writer's finished product. Nor was it possible to signal to the writers that what was wanted in each exercise was a noncompounded sentence. In other words, though the exercises were designed to measure a writer's competence to subordinate sentential statements by means of noun-modifier transformations,

to the exclusion of alternative strategies such as intersentential compounding and intrasentential conjoining, no terms were available with which to communicate this message explicitly to the writers. In other words, the SC assessment did not yield a pure measure of what writers can do, since conditions of practicability required use of an exercise format open enough to allow some of what they choose to do to creep in as well.

D. Scoring Procedures

Results of the SC exercises were categorized by the same team of trained English teachers who performed the syntax and mechanics analysis of the essays written in the 1978-79 writing assessment (see NAEP 1980a, b, c). The scoring procedure, developed by NAEP staff members and professional consultants allows for classification of responses in terms of overall syntactic and semantic structure and in terms of the sequence of sentence-combining transformations. Here is the scheme in detail:

I. Overall syntactic and semantic structure: Responses were placed within each of the following four categories:

1. Lexical content: Are the nouns, verbs and adjectives found in the given sentences preserved in the responses? This factor was scored as follows:
   a. Content preserved without addition or omission
   b. Some content added
   c. Some content omitted
   d. Some content added or some omitted

2. Syntactic relations: Are the grammatical relationships found in the given sentences preserved in the response? This factor was scored:
   a. Syntactic relations synonymous with those of the given sentences
   b. Syntactic relations are different from those of the given sentences
   c. Syntactic relations introduce ambiguity into the response
   d. Syntactic relations introduce ineptitudes into the response -- awkward and inappropriate constructions, faulty parallelism, misplaced modifier, agreement error, and so on.
   Examples:
   The rope that was limp hung from the tree branch was a clue to the mystery.
   Forest fires often destroy lives and property by careless people who drop lighted cigarettes.
   They waited for a bus at the corner which was twenty minutes late.

3. Inversions: Is the main clause of the response in normal order or inverted order (passive, "there" or "it" expletive inversion or cleft inversion)? This factor was scored:
   a. Normal order
   b. Inverted order

4. Number of T-units (independent clauses): How many T-units (i.e., independent clauses, or alternatively, main clauses plus all the subordinate structures each happens to contain) comprise the response? This factor was scored:
   a. All one T-unit
   b. Two T-units
   c. Etc., up to seven T-units

II. Sentence-combining transformations: Responses were categorized according to the sequence of sentence-combining transformations utilized in their formation. Four categories of transformations were used, exactly the same as those employed in the syntax analysis of the essay exercises.

1. Adjectival embeddings
   Prenoun modifiers (words or hyphenated phrases)
   Postnoun modifying words and phrases (includes adjective prepositional phrases and appositions)
   Postnoun relative clauses
   Postnoun nonreduced relative clauses (should have been reduced/should have been scorable as an 11 or 12)

2. Nominal embeddings
   "One-word" (uncomplemented) gerunds or infinitives
   Nominal phrases (gerunds or infinitives)
   Nominal clauses (fact "that" or question clauses)

3. Adverbial embeddings
   Single-word adverbs and adverbial prepositional phrases
   Verbal phrases (infinitive phrases, gerundives following time, manner, etc.)
   Adverbial clauses (place, time, manner, reason, purpose, condition, concession, etc.)

4. Conjoinings
   Conjoined verbs, predicate phrases, noun phrases
   Participial conjoinings, nominative absolutes
   Coordinate compounding (any instance of two or more T-units compounded by coordinating conjunctions and, plus, for, but, yet, etc.)

Conjunctive adverbial compounding (any instance of two or more T-units)
Fused, comma splice, run-ons

Detailed criteria for identifying the syntactic structures named in the above list are given in the NAEP document no. 10-W-50, Guidelines for Three Ways of Evaluating Writing: Syntax, Cohesion and Mechanics, by Ina V.S. Mullis and John C. Mellon (1980).

Summarizing, the strategy for scoring the SC exercises was first to differentiate those that preserved the lexical and grammatical meanings of the given exercises from those that altered these meanings, and then to identify the number of T-units contained in each. Finally, responses were analyzed in terms of the pattern of sentence-combining transformations observed in each. In consequence, a "correct" response to a given exercise was defined as any combination, whether inverted or noninverted, that preserved lexical and grammatical synonymy and consisted of a single T-unit (independent clause).

Responses that were blank, illegible, illiterate, unrelated to the given exercise, fragments, copies of the given sentences or copies of the given sentences conjoined with the same conjunction were not ratable using this scoring procedure. These types of responses ranged from 21% to 36% at age 9, 4% to 14% at age 13 and 3% to 6% at age 17. (See Table 2-1.) These responses are included under the heading "Nonratable" in the various tables of this report.

E. The Analysis and Data Presentation

National Assessment reports the performance of groups of students, not individuals. Because the numbers and percentages presented in this report are based upon samples, they are necessarily estimates, not definitive measures of national populations. They are, of course, the best estimates, but they are subject to the qualification that a certain amount of measurement and nonmeasurement error creeps into even the best estimates. Thus, for example, the figure 20% is really 20% plus or minus a certain (usually small) margin of error.

National Assessment computes standard errors that estimate the sampling error and other random error associated with the assessment of a specific item. NAEP has adhered to the standard convention whereby differences between statistics are designated as statistically significant only if the differences are at least twice as large as their standard errors. Differences this large would occur by chance in fewer than 5% of all possible replications of the sampling, data collection and scoring

procedures for any particular age group or reporting group. If a national figure was 20% and if the standard error of the female percentage was .5 points, 22% would be "significantly" (in the statistical sense) different from 20% because it is more than twice the standard error away from 20%. But if the percentage for females was 20.5%, it would not be at least twice the standard error of the change estimate away, so it would not be termed a statistically significant difference.

It is important, however, to distinguish statistical significance from educational significance. A difference of 3 or 4 points between group and national performance might be statistically significant but too small to merit serious educational concern. One can also imagine a situation in which many changes are negative but no one of them is statistically significant; it could be that the overall pattern of negative changes has educational significance. Readers must decide for themselves how important particular changes or differences are in the real world, for statistical conventions can aid, but not replace, good judgment.

CHAPTER 2

NATIONAL RESULTS AND DISCUSSION

The SC Exercises

Fourteen SC tasks or exercises are included in this report. Of these, 8 were administered to all three age groups, 4 to the 13- and 17-year-olds only and 2 to the 9-year-olds only. Shown below are the 14 exercises, listed in order of difficulty as determined by the percentage of writers producing acceptable combinations, and identified by a shorthand title used throughout the balance of this report. Also shown is the most frequently observed acceptable combination. The percentage of all writers who wrote that combination appears in appendix Table A-2.

1. Exercise title: "Coat"

   As given: Bill's coat was in the closet.
   The coat was new.
   It was leather.

   Most frequent combination:
   Bill's new leather coat was in the closet.

2. Exercise title: "Clown"

   As given: The clown was smoking a cigar.
   The clown was jolly.
   The cigar was fat.

   Most frequent combination:
   The jolly clown was smoking a fat cigar.

3. Exercise title: "Lemonade"

   As given: The boys drank the lemonade.
   The boys were barefoot.
   The lemonade was cold.

   Most frequent combination:
   The barefoot boys drank the cold lemonade.

4. Exercise title: "Cries"

As given: Her cries were lost in the storm.
Her cries were thin.
Her cries were small.

Most frequent combination:
Her thin small cries were lost in the storm.

5. Exercise title: "Rope"

As given: A rope was the clue to the mystery.
The rope was twisted.
The rope was hanging from a tree branch.

Most frequent combination:
A twisted rope hanging from a tree branch was the clue to the mystery.

6. Exercise title: "Magician"

As given: John knows a magician.
The magician is clever.
The magician can make an elephant disappear.

Most frequent combination:
John knows a clever magician who can make an elephant disappear.

7. Exercise title: "Troops"

As given: The captain took care of his troops.
The captain was strong.
He was fearless.
The troops were tired.
The troops were hungry.

Most frequent combination:
The strong fearless captain took care of his tired hungry troops.

8. Exercise title: "Bubble"

As given: The plants are kept dry by a bubble.
The bubble is large.
The bubble is plastic.
The bubble covers the entire garden.

Most frequent combination:
The plants are kept dry by a large plastic bubble that covers the entire garden.

9. Execise title: "Forest Fires"

As given: Careless people often cause forest fires.
Careless people tend to drop lighted cigarettes.
Forest fires can destroy lives and property.

Most frequent combination:
Careless people who tend to drop lighted cigarettes often cause forest fires that can destroy lives and property.

10. Exercise title: "Guard"

As given: A guard kept the children from touching the animals.
The guard was bored.
The guard was at the doorway.
The animals were dusty.
The animals were stuffed.
The animals were in the museum display.

Most frequent combination:
A bored guard at the doorway kept the children from touching the dusty, stuffed animals in the museum display.

11. Exercise title: "Pebbles"

As given: The pebbles marked the path to a kingdom.
The pebbles were shiny.
The pebbles were yellow.
The pebbles were gleaming like cats' eyes.
The kingdom was magic.
The kingdom was underground.
The kingdom was ruled by a wizard.

Most frequent combination:
The shiny yellow pebbles gleaming like cats' eyes marked the path to a magic underground kingdom ruled by a wizard.

12. Exercise title: "Hikers"

As given: The hikers tramped along the path.
The path was steep.
It was narrow.
It was rocky.
It curved upward toward the mountain top.
The mountain top appeared ahead through the clouds.

Most frequent combination:

The hikers tramped along the steep narrow rocky path that curved upward toward the mountain top that appeared ahead through the clouds.

13. Exercise title: "Lookout"

As given: The lookout was frightened.
He was clinging to the mast.
He realized the tidal wave would swamp the ship.
The wave would send it plunging to the depths.

Most frequent combination:
The frightened lookout clinging to the mast realized the tidal wave would swamp the ship and send it plunging to the depths.

14. Exercise title: "Bus"

As given: The people were standing on the corner.
Their hands were cold.
Their hands were dangling by their sides.
They were waiting for a bus.
The bus was already twenty minutes late.

Most frequent combination:
The people standing on the corner with their cold hands dangling by their sides were waiting for a bus that was already twenty minutes late.

It is important to remember, as one examines these 14 SC exercises, that the combinations shown above are by no means the only acceptable solutions to the given tasks, but are merely the patterns most frequently used by the 2,500-2,800 students at each age who wrote each exerise. As is shown in appendix Table A-1, the diversity of acceptable patterns actually produced is quite large. Of interest also is that in every case the most frequently observed pattern was the same for all ages and proved acceptable in terms of grammatical and semantic conformity to the basic sentences of the exercise as given.

Table 2-1 presents the national results from the 14 SC exercises for all ages at which they were administered. The exercises are arranged in ascending order of difficulty as defined by the average correctness percentage registered by the 13-year-olds and 17-year-olds. To reiterate, the term "correct" refers throughout this report to combinations one T-unit in length that are free from grammatical ineptitudes and preserve the lexical content and grammatical relations of the given sentences. Note, however, that Table 2-1 also shows responses two T-units long that are otherwise correct.

TABLE 2-1. National Results: Percentages of Students Writing Correct, Incorrect and Nonratable Combinations#

Exercise Title	Age	% Ratable Responses			% Nonratable Responses
		Correct and 1 T-Unit	2 T-Units, Otherwise Correct	Incorrect	
Coat	9	38	6	35	21
Clown	9	40	4	30	26
Lemonade	9	34	8	28	30
	13	77	4	13	7
	17	91	2	6	2
Cries	9	28	7	33	32
	13	66	4	22	7
	17	84	2	12	2
Rope	9	16	7	48	29
	13	60	7	28	5
	17	80	4	14	2
Magician	9	16	7	46	31
	13	52	7	34	7
	17	76	4	18	3
Troops	9	20	4	50	27
	13	52	6	38	4
	17	70	4	24	2
Bubble	9	18	4	48	30
	13	49	3	40	8
	17	65	2	30	3
Forest Fires	13	40	16	37	6
	17	62	9	26	3
Guard	9	6	2	56	36
	13	32	4	54	10
	17	61	3	32	3
Pebbles	9	3	0	63	33
	13	21	2	69	8
	17	48	3	47	2
Hikers	13	23	4	66	7
	17	44	5	50	2

TABLE 2-1 Continued.

Exercise Title	Age	% Ratable Responses			% Nonratable Responses
		Correct and 1 T-Unit	2 T-Units, Otherwise Correct	Incorrect	
Lookout	13	20	10	56	14
	17	44	12	39	5
Bus	13	12	6	75	8
	17	33	7	57	3

#Figures may not total due to rounding.

Almost certainly, much of the difficulty of any SC exercise depends upon the number and kind of transformational operations each one requires. An exercise analysis along these lines is given later in this report (Table 2-4). Here at the outset, perhaps the most obvious point to note is that, on every exercise, older students do better -- a not unexpected result, which further confirms the age-relatedness of syntactic skills observed by practically all researchers to date.

Second, in the case of the youngest writers, the nonratable responses tell us that between one-fourth and one-third of all 9-year-olds cannot perform even simple SC tasks. Almost certainly, however, their inability does not stem from nonacquisition of the grammatical operations in question but rather from naturally late developmental schedules in any or all of the following areas: motor skills in handwriting, reading and spelling ability, confidence in undertaking tasks requiring conscious problem solving in language and ability to "hold" language in short-term memory while inscribing it.

On the two exercises most difficult for the 9-year-olds, however, "Guard" and "Pebbles," the nonratable totals are only a little higher than in the other exercises, while the correctness totals plummet almost to zero. Here we do see true lack of ability, whether to perform certain syntactic operations or to process information cognitively. Lack of ability does not mean, of course, that anything is amiss with these 9-year-olds. We have merely set them a task, in the case of exercises like "Guard" and "Pebbles," that they are not as yet mature enough to handle successfully.

Table 2-1 also reports the percentage of correct responses two T-units in length, the result either of compounding independent clauses or of joining them by conjunctive adverbs or fused punctuation. As noted earlier, no way could be found that was nontechnical yet comprehensible in which to word instructions calling for single-T-unit responses. Experience with pedagogical SC indicates that a two-T-unit response often represents a writer's half-conscious "fall-back" strategy when the given exercise proves particularly difficult -- that is, when it requires real virtuosity to transform all the given sentences into subordinate structures under the dominance of a single main-clause predicate. (Some pedagogical exercises, of course, are given in forms specifically cued for multi-T-unit responses.) As expected, the 17-year-olds resorted least to this fall-back strategy, except in the case of "Forest Fires," which required two nonreducible relative clauses, and the two most difficult exercises, "Lookout" and "Bus."

On the other hand, except for "Guard" and "Pebbles," which in effect were beyond the reach of the 9-year-olds in any form, the 9-year-olds appear to show no tendency whatever toward wanton overcompounding, a feature often found in syntactic analysis of their free writing. In general, despite large gaps between each pair of ages on both easy and difficult items in the one-T-unit correctness percentages, the two-T-unit percentages of the 9-year-olds and 13-year-olds are quite alike on the easy items, and those of the 13-year-olds and 17-year-olds are similarly alike on the difficult items. One would have thought that in each case, more writers at the lower age would resort to a multi-T-unit fall-back strategy; the fact that they do not remains unexplained.

Are the overall correctness percentages as high as one might predict or hope for? Perhaps not. NAEP's 1978-79 assessment represents the first attempt anywhere to gather normative data on SC skills, and there is literally nothing in the way of existing information with which to compare these results. Some might conclude, because SC is mostly unused as a type of test item, and in all probability is known as a pedagogical activity by fewer than half of the students sampled in the assessment, that the observed correctness totals are not surprisingly low. Others, however, will point out that the exercises were presented in the form of simple editing tasks that should have seemed understandable to all students, and that the correctness totals are indeed low.

On further reflection, however, while it is true that the three-sentence exercises might have seemed like ordinary editing tasks, there is no way that lists of six or seven sentences, such as those comprising "Guard" and "Pebbles," for example, could be taken for anyone's first-draft writing. Since a given student

worked no more than three exercises, the last invariably being the most difficult, it is likely that many writers experienced a kind of "Oops!" reaction upon suddenly finding the third, and in some instances the second, exercise not at all a matter of simple editing, but rather a tough problem requiring difficult syntactic manipulation and the processing of a great deal of information. Many may not have recovered from the "Oops!" in time to do their best work. Relevant here is the testimony of specialists in pedagogical SC, who invariably report that although nearly all students can fairly quickly "get the hang" of combining anomalous-seeming lists of sentences, they definitely require at least one substantial training session involving a number of practice opportunities in which to do so. At present, unless and until pedagogical SC becomes more widespread, it seems appropriate to conclude that use of SC exercises as test or assessment items without prior orientation and/or practice may yield artificially low correctness totals.

Turn now to the students whose combinations were ratable but incorrect, that is, who either cast their responses in multiple T-units, or altered the given lexical content or syntactic relations, or did some combination of the three. Tables 2-2 and 2-3 show the percentage results on these factors. In looking at these tables, one should be aware that the correctness percentages shown in Table 2-1 are contained within the "One T-Unit" percentages of Table 2-2, and the two "None" percentages of Table 2-3, but in no case are they identical, since at no age did all writers who wrote only one T-unit also succeed in preserving both lexical content and syntactic relations.

Table 2-2 tends to go against the conclusion tentatively advanced on the basis of Table 2-1 results, since it shows that among 9-year-olds' writing ratable responses, as opposed just to those writing correct responses, the tendency to write multiple T-unit combinations (that is, not to embed) is quite pronounced after all, and relatively uniform regardless of the difficulty of the exercise. In fact, with "Troops," "Guard" and "Pebbles," which required the combining of five, six and seven sentences, respectively, more 9-year-olds wrote multiple T-units than wrote single T-units -- clear evidence of their inability to retain and process information while at the same time performing embedding operations.

TABLE 2-2. Total Percentage of Ratable Responses Divided According to the Number of T-Units Comprising Each Response

Exercise Title	Age	Total % Ratable Responses	Total % Ratable Responses by T-Unit Count#					Number of Sentences in the Exercise
			1 T-Unit	2 T-Units	3 T-Units	4 T-Units	5 or More T-Units	
Coat	9	79	52	15	12	0	0	3
Clown	9	74	58	10	5	0	1	3
Lemonade	9	70	46	14	9	0	0	3
	13	93	86	6	1	0	0	
	17	98	95	2	1	0	0	
Cries	9	68	53	11	3	0	2	3
	13	93	85	6	1	0	1	
	17	98	95	2	0	0	1	
Rope	9	71	38	19	13	0	1	3
	13	95	80	11	3	0	1	
	17	98	92	5	1	0	0	
Magician	9	69	33	20	15	0	0	3
	13	93	73	15	4	0	1	
	17	97	88	8	1	0	0	
Troops	9	73	34	16	16	7	1	5
	13	96	66	15	11	3	1	
	17	98	81	10	6	1	0	
Bubble	9	70	44	14	7	4	1	4
	13	92	79	8	2	1	2	
	17	97	91	4	1	0	1	
Forest Fires	13	94	64	25	4	0	1	3
	17	97	81	13	2	0	0	

TABLE 2-2 Continued.

Exercise Title	Age	Total % Ratable Responses	Total % Ratable Responses by T-Unit Count*					Number of Sentences in the Exercise
			1 T-Unit	2 T-Units	3 T-Units	4 T-Units	5 or More T-Units	
Guard	9	64	29	16	8	11	1	6
	13	90	64	14	6	4	2	
	17	97	82	10	3	1	1	
Pebbles	9	67	24	13	9	20	1	7
	13	92	62	16	7	4	3	
	17	98	77	13	4	2	2	
Hikers	13	93	60	18	9	5	1	6
	17	98	76	14	5	2	0	
Lookout	13	86	45	26	10	4	1	4
	17	95	65	21	6	2	0	
Bus	13	92	46	24	14	7	3	5
	17	97	65	18	9	2	3	

*Figures may not total due to rounding.

TABLE 2-3. Total Percentage of Ratable Responses Divided According to Alterations in Given Lexical Content and Syntactic Relations

Exercise Title	Age	Total % Ratable Responses	% Lexical Alterations				% Syntactic Alterations#		
			None	Content Added	Content Omitted	Added & Omitted	None	Well-Formed Alterations	Inept Alterations
Coat	9	79	56	4	16	3	70	1	9
Clown	9	74	55	5	12	2	59	10	5
Lemonade	9	70	56	3	9	2	60	2	8
	13	93	84	1	8	1	87	1	5
	17	98	95	1	2	0	95	0	3
Cries	9	68	48	2	17	2	47	4	17
	13	93	82	1	9	1	76	3	13
	17	98	93	1	4	0	88	2	8
Rope	9	71	36	2	30	3	54	5	12
	13	95	77	1	16	1	82	2	11
	17	98	92	1	5	0	89	3	6
Magician	9	69	47	3	16	3	37	6	26
	13	93	79	0	12	1	67	4	22
	17	97	92	0	5	0	82	2	13
Troops	9	73	43	2	25	4	53	9	11
	13	96	77	1	17	1	76	11	9
	17	98	85	2	11	1	86	6	6
Bubble	9	70	34	1	30	4	47	4	18
	13	92	62	2	26	2	72	4	16
	17	97	74	1	21	1	83	4	10
Forest Fires	13	94	73	3	15	2	71	2	20
	17	97	84	2	9	1	81	1	16

TABLE 2-3 Continued.

Exercise Title	Age	Total % Ratable Responses	% Lexical Alterations				% Syntactic Alterations#		
			None	Content Added	Content Omitted	Added & Omitted	None	Well-Formed Alterations	Inept Alteration
Guard	9	64	19	2	37	5	36	9	18
	13	90	47	2	38	4	60	7	23
	17	97	72	2	21	2	80	4	12
Pebbles	9	67	15	1	42	8	31	12	23
	13	92	37	1	50	4	49	14	29
	17	98	66	1	28	3	67	10	21
Hikers	13	93	44	1	44	4	59	12	22
	17	98	68	2	26	3	69	11	14
Lookout	13	86	43	1	38	4	56	6	25
	17	95	67	3	22	2	76	1	18
Bus	13	92	43	1	46	2	43	9	40
	17	97	67	2	26	2	60	6	31

#Figures may not total due to rounding.

Looking across all three ages, one sees a uniform reduction in the percentage of multi-T-unit responses on any given exercise, except that the more difficult the task, the greater the number of multi-T-units at each age. Experience with performance of older students first introduced to pedagogical SC indicates that unwanted multi-T-unit responses soon disappear, whereas younger children cannot readily be taught to embed very much information into single T-units, but must "grow into" the ability at their own pace, over time. Again, one may speculate that a dramatic reduction of multi-T-unit responses would be observed in many cases among the 13-year-olds, and across the board among the 17-year-olds, in an assessment setting, were the students given even a brief advance period of instruction and practice.

In Table 2-3, one sees that the observed lexical alterations were almost exclusively matters of content omitted from the given exercise, and the more difficult the exercise, the greater the percentage of writers who omitted content. Only in the case of "Forest Fires," whose difficulty stems more from the syntactic requirement of forming two unreduced relative clauses than from the inclusion of lexical content from increasing numbers of given sentences, is there a really marked disparity between difficulty level of the exercise and percentage of writers omitting content.

Syntactic alterations, also shown in Table 2-3, were, as one might have expected, mostly cases of ineptitude, although there were also significant numbers of well-formed alterations, especially in exercises containing the largest number of sentences to be combined. These results are not surprising. The ineptitudes may result simply from the writers' failure to reread their productions and "listen" for grammaticalness, then revise where necessary. Or they may indicate that the writers had gotten themselves into syntactic waters over their heads, so to speak, where they could not, by rereading and reflecting upon what they had written, discern whether their structures were well-formed or not. Examination of the students' response sheets reveals very little in the way of revision of any kind, and their unwillingness and/or irability to test and revise their productions undoubtedly accounts for the rather high instance of syntactic ineptitude. Among the 9-year-olds, such ineptitude is to be expected and tolerated, since it will diminish in due time. But with the older writers, where the median percentage of 13-year-olds registering ineptitudes on the more difficult half of the exercises was 23% and that of the 17-year-olds was 16%, one wonders how much of the ineptitude comes from lack of trying and how much stems in fact from lack of training.

Grammatical Analysis of the Exercises

At this point, it is appropriate to look more closely at the particular structures involved in each exercise. Table 2-4 presents this information, ranking the exercises in ascending order of difficulty. In the easier exercises, the primary indicator of difficulty seems to be number of basic sentences to be combined, so long as their targeted surface structures are simple word or phrase modifiers, as is the case in the first eight exercises. In "Forest Fires," as noted above, the presence of two nonreducible clauses accounts for the difficulty of this three-sentence exercise, a difficulty possibly compounded by a restrictive/nonrestrictive punctuation dilemma that remains unsolvable so long as the sentence appears out of a discourse context. "Guard" and "Pebbles" are difficult in that they require a large number of modifiers to be attached to two nouns rather than one; that is, they require the building of two complex noun phrases within one sentence, not merely one.

"Hikers," "Lookout" and "Bus" are the most difficult of the exercises, at least partly because all three require modifications at the "second level" of embedding; they require that some noun within a modifying structure itself be modified by yet another modifying structure, a situation that linguists refer to as "nested" or "recursive" embedding. In "Hikers," for example, "path" in the main clause is modified by four first-level embeddings, after which "mountain top," a noun contained within one of these first-level structures, must itself be modified by a second-level embedding maximally reducible to the phrse "appearing ahead through the clouds." Syntactic maneuvers like this are easy enough, even routine, for mature and practiced writers, but they are quite difficult for most young people even through the high school and college freshman years.

The last three exercises are syntactically challenging for still other reasons. "Hikers," for example, is the only exercise requiring the attachment of four modifying structures to a single noun. "Lookout" is the only exercise that includes nominalization, but the nominal is already contained in the exercise, in the third sentence given, whose pronoun subject "he" must therefore be replaced by the noun "lookout." Moreover, the fourth sentence of "Lookout" is best rendered by the participial conjoining, "sending it plunging...," a form typically mastered by only the most proficient writers, and here chosen by only 72 of the nearly 2,800 17-year-olds who wrote the exercise.

TABLE 2-c. Syntactic Features of the 14 SC Exercises

Exercise Title	% Correct, 13- & 17-Year-Olds Averaged	Number of Sentences Included	Furthest Transformational Reductions Predicted			Modifications Per Noun		Embeddings at Each Depth Level	
			Words	Phrases	Clauses	First Noun	Second Noun	First Level	Second Level
Coat	—	3	2	-	-	2	-	2	-
Clown	—	3	2	-	-	2	-	2	-
Lemonade	84	3	2	-	-	1	1	2	-
Cries	75	3	2	-	-	2	-	2	-
Rope	70	3	1	1	-	2	-	2	-
Magician	64	3	1	-	1	2	-	2	-
Troops	61	5	4	-	-	2	2	4	-
Bubble	57	4	2	1	-	3	-	2	-
Forest Fires	51	3	-	-	2	1	1	2	-
Guard	47	6	3	2	-	2	3	5	-
Pebbles	34	7	4	2	-	3	3	6	-
Hikers	34	6	3	2	-	4	1	4	1
Lookout	32	4	1	2	1	2	-	3	1
Bus	22	5	1	2	1	2	1	2	2

"Bus," finally, invited two kinds of participial constructions, one a nominative absolute ("the people were standing on the corner, their cold hands dangling..."), and the other a participial conjoining containing an unreduced relative clause ("waiting for a bus that was..."). Not very surprisingly, perhaps, only 16 of 2,800 17-year-olds used this pattern. The most frequent pattern, which introduced a "with their cold hands..." adverbial phrase, was observed in 156 instances. Overall, "Bus" provided opportunities for by far the greatest number of correct patterns observed, 206 among the 17-year-olds. The exercise is constituted in such a manner that any one of the given sentences may become the main clause, so that possible grammatical relationships among the given sentences have to be thought about in an overt "problem-solving" mode, rather than pursued more or less routinely, as in the easier exercises. Its validity as an SC task is therefore high pedagogically, yet it will always seem difficult cognitively, and will yield low correctness totals, as an assessment item.

Table 2-5, in turn, shows something of the astounding diversity of correct patterns the students were able to discover. Many of these, while not ungrammatical (that is, not scored as an ineptitude), were what one might term, for want of a better label, "stylistically weak." For example, in "Coat," despite the fact that 95% of the children who gave a correct response wrote "Bill's new leather coat was in the closet," another pattern was "Bill's new coat was leather and was in the closet." Adult writers would agree that the conjoined predicates here, while certainly not ungrammatical, are not really in parallel one with the other, because of their differing content.

In the final six exercises, the difficult ones, the 17-year-olds show their virtuosity by discovering many more correct patterns than the younger writers find. But in the easier exercises, the "pull" of the exercise toward the obvious pattern is so strong for the 17-year-olds that they use fewer patterns than the 13-year-olds. In "Bubble," for example, one of the active-voice inversion patterns used by several 13-year-olds but no 17-year-olds was "a large plastic bubble covers the entire garden so that the plants are kept dry." Again, this is neither ungrammatical nor particularly offensive, yet one is somehow happy to see the 17-year-olds avoid it. A general observation is, the larger the number of correct patterns, if coupled with few instances of any one pattern, the more difficult the exercise, as Table 2-5 consistently shows. In other words, the difficulty of an exercise is determined in part by the degree of diversity of possible responses to it, and not merely by the number of sentences it contains, or the type of transformations required.

TABLE 2-5. Diversity of Syntactic Structures in 14 SC Exercises

Exercise Title	Number of Total Correct Patterns Observed			Percentages of Total Correct Combinations Instanced by Most Frequently Observed Pattern		
	Age 9	Age 13	Age 17	Age 9	Age 13	Age 17
Coat	10	—	—	95	—	—
Clown	23	—	—	88	—	—
Lemonade	23	23	19	55	77	82
Cries	20	27	23	52	78	87
Rope	22	35	23	63	69	73
Magician	13	31	25	74	84	93
Troops	17	34	37	87	90	88
Bubble	24	30	20	74	72	74
Forest Fires	—	44	54	—	45	48
Guard	36	71	139	53	64	61
Pebbles	28	71	80	18	24	26
Hikers	—	32	38	—	81	84
Lookout	—	59	98	—	23	24
Bus	—	139	206	—	14	18

An idea of the sentence patterns most frequently preferred may be gained from appendix Table A-1, which shows the correct responses used by 10 or more writers at each age, ranked by frequency of occurrence, together with the percentage of students writing correct responses who used the particular pattern. In other words, Table A-1 orders the patterns of correct responses and shows what percentage of correct responses can be attributed to each pattern.

Recalling the exercise directions -- "Each set can be improved by combining the given sentences into <u>one</u> sentence that

says the same thing" -- one sees that all a writer had to do to satisfy the request was to recast the given sentences as a single sentence. Nothing was said about the form of that sentence. Still, classroom observation of pedagogical SC shows that some students seem naturally to craft their sentences in special ways, with varying degrees of artistry, while others do not. Even apart from a discourse context, sentences containing such things as nonrestrictive elements, parallel constructions, nominative absolutes, participial conjoinings, introductory modifiers, nonnormal word orders, and so on, when well chosen, strike most persons as fine or artistic writing, however they may characterize it. Few would disagree, for instance, that a fine-writing response to "Cries" might be "small and thin, her cries were lost in the storm." Table A-1 fails to show this pattern because it was not observed 10 or more times at any age. In fact, exactly 5 students out of 2,766 17-year-olds did produce this response, as did 3 13-year-olds. And Table A-1 does show that the equally artistic form, "Her cries, thin and small, were lost in the storm," was the third-ranked pattern among the 17-year olds, used in 2% (35) of the correct responses. But in general, Table A-1 supports the conclusion that "fine writing" responses were rare in the SC assessment. Whether this rarity indicates lack of ability or merely lack of inclination cannot be determined, and some of the exercises hardly lend themselves to artistry in any case, no matter how clever the writer.

Table A-1 may also be examined from the opposite perspective, that of stylistic weakness, infelicity, nonungrammatical "awkwardness," or whatever one likes to call it. Whereas one cannot really fault young writers for failing to display conscious artistry, one might hope on the other hand that there would be fewer instances of linguistically "dead ears" than Table A-1 seems to indicate. Of course, some of the problems are known to be transitory, that is, they disappear with maturity and writing practice. Failure to reduce relative clauses beginning with a relative pronoun followed by a form of "be" is perhaps the most common illustration. In "Rope," for example, the second most frequent pattern was "a twisted rope that was hanging from a tree branch was a clue to the mystery." The words "that was" are syntactically deletable, perform no semantic or stylistic function whatever, nor would they act as a device of cohesion in any discourse context in which the sentence might occur. They are pure deadwood, almost sure to be pruned (or never written in the first place) by careful adult writers. Younger writers, however, have first to learn to write the full relative clause, and only afterwards do they learn to make appropriate and mostly automatic relative-pronoun-plus-"be" reductions. This phenomenon has traditionally been observed by secondary-school writing teachers, and is commonly seen in pedagogical SC as well. Other instances of unreduced relative clauses are in "Hikers," where the second most frequent pattern among the 13-year-olds included "the...path that

was curving upward..."; in "Lookout," where one sees "...the lookout who was clinging to the mast..."; in "Guard," "...guard that was at the doorway..." and "...animals that were in the museum display"; in "Pebbles," "...pebbles that were gleaming like cat's eyes...," and so on.

Also noticeable is the tendency of some writers to prefer coordination to subordination in the combining of sentences. Doing so permits single-word modifiers, for example, to retain their status as surface-structure predictions without becoming separate sentences, as in "Hikers," "The path the hikers tramped along was steep, narrow and rocky, and curved upward..." or in "Troops," "The captain was strong and fearless and took care of his tired hungry troops." In "Lookout," the most frequent pattern contains "...the tidal wave would swamp the ship and send it...," whereas a writer conscious of the interplay of temporal and causal nuances in consecutive predictions might prefer, as many 17-year-olds but fewer 13-year-olds did, the participial conjoining "...the tidal wave would swamp the ship, sending it plunging to the depths."

Some writers use explicit temporal connectives in the conjoining of sentences, often attenuating the resultant structure. In "Lookout," once again, the second most frequent pattern among the 17-year-olds, and the most often used among the 13-year-olds, contained "...clung to the mast as he realized... ." While the construction is certainly not wrong in any clearcut way, many writers would agree that the temporal "as" works better with "clung," a continuous physical action, than with "realized," a noncontinuous mental action, and thus might prefer "frightened as he clung to the mast, the lookout realized...," or "the frightened lookout clung to the mast, realizing... ." Similarly, in "Hikers" the fourth most frequent pattern among the 17-year-olds contains "the hikers tramped along the...path as it curved upward...," an ill-considered and awkward use of "as," though impossible to label as a definite error.

Other exercises showed problems seeming to arise from a writer's choice of main clause statement, particularly "Forest Fires." One pattern used by 10% of the 13-year-olds who wrote correct responses avoids entirely the question of choosing main clause, and simply conjoins the three predicates: "Careless people tend to drop lighted cigarettes, often cause forest fires, and destroy lives and property." Underlying this exercise as given is an explicit causal sequence: people drop cigarettes, cigarettes cause fires, fires destroy lives and property. The wording of the exercise, however, made "people" not "cigarettes" the cause of the fires, by a kind of logical transformation. Writers of the pattern just quoted transformed the logic a step further, making "people" the agent of the whole business -- the dropping, the forest fires and the destruction of lives and property. A logically exact rendering of the exercise would be, "Careless people tend to drop

lighted cigarettes which often cause forest fires that can destroy lives and property." Students who wrote the foregoing were not scored as having deviated from the given lexical or syntactic relations, since they were in fact clarifying the underlying logical sequence. But the fact that the pattern "careless people who tend to drop lighted cigarettes often cause forest fires that can destroy lives and property" yields the same scoring pattern as the foregoing (two relative clauses) means that these two patterns are conflated as the first-ranked choice shown in Table A-1 for "Forest Fires."

Another kind of problem observed in correct responses is that the line of modification within a pattern may be unclear. In one of the patterns found in "Forest Fires," one sees what grammar books sometimes call a "squinting modifier," "Careless people often cause forest fires by dropping lighted cigarettes, which can destroy lives and property." The syntax says it is the cigarettes that destroy lives and property, but logic tells us it is the forest fires caused by the dropping of the cigarettes that do the destroying. In other words, the modifier seems to be looking one place but actually is looking another.

Another pattern shows what the handbooks term "remote modification." In the pattern "people often cause forest fires that can destroy lives and property by carelessly dropping lighted cigarettes," there is too much between "people" and "by dropping" that the "by dropping..." phrase might seem to modify; hence, one feels a definite lack of smoothness in the modification, that is, an uncertainty as to the implicit prediction it signals.

Finally, in "Bus" one sees patterns resulting from four different choices of main-clause prediction: "people were waiting...," "people stood/were standing...," "there were people dangling their hands..." and "people had cold hands... ." Only the last of the given sentences was not chosen as main clause, since the pattern it would have yielded, while perfectly grammatical, is rather difficult to perform: "The bus for which the people standing on the corner and dangling their cold hands by their sides were waiting was already twenty minutes late." Interestingly, despite the fact that many persons mistakenly assume the nominative absolute old-fashioned or pretentious, one of the crispest and most straightforward patterns of "Bus" contained not one but two absolute noun phrases: "The people stood on the corner, their cold hands dangling by their sides, the bus they were waiting for almost twenty minutes late."

In general, unlike exercises such as "Pebbles," which contain more constituent sentences but offer fewer options for differing patterns, exercises like "Bus" can be viewed not only as tests of writers' ability to combine into single statements sentences whose interrelationships are not syntactically obvious, but also as

screening devices separating writers into three groups: one, those who do so smoothly but not in any particularly artistic or memorable way (in "Bus" and indeed all the other exercises, this was always the first-ranked pattern); two, those who do so artistically by means of syntax typifying fine writing; and three, those whose combinations, while error-free technically, are clumsy and awry. National Assessment's scoring procedure in the present assessment attempted no such three-way separation of writers, but it might well do so in the future. What is clear from the present assessment, of course, is that the percentage of "fine writing" would be very low indeed, even among the 17-year-olds. Still, from the point of view of their being a reliable and efficient device allowing objective comparisons among different writers, it is interesting to speculate that specifically designed SC exercises might actually be more effective than free writing as a means of identifying young people with latent talent for artistic written expression.

CHAPTER 3

RELATIONSHIP OF SENTENCE-COMBINING SKILL TO OTHER VARIABLES

National Subgroups

Table 3-1 shows all instances in which the percentages of correct responses for a subgroup were different from the nation, either lower or higher, to a statistically significant degree.[1] It was hypothesized at the outset, on the basis of inferences from developmental research on normal growth of syntactic fluency, as well as from studies of the effectiveness of pedagogical SC, that SC ability would vary among individuals much as do other, more familiar academic abilities assessed by National Assessment. Table 3-1 bears out the hypothesis. Girls generally outperformed boys at all ages, 26 times in 34 possibilities. Black students performed significantly below the nation on every exercise at every age, while white students performed above the nation. Similarly, students whose parents lack high school diplomas tended to achieve below the nation by rather large amounts, while students whose parents had some postsecondary education performed above the nation. The situation is much the same for students from disadvantaged- and advantaged-urban areas, respectively.

On the five most difficult tasks ("Guard" through "Bus") the 13-year-olds in the Northeast significantly outperformed the nation in every case. This is a striking regularity, and perhaps betokens wider use of pedagogical SC in middle-school and junior-high curricula in the schools of the Northeastern region during the years immediately preceding the 1978-79 assessment. This surmise becomes a virtual certainty if one grants the assumption that acceptable performance on difficult SC tasks is highly dependent upon prior in-school practice on such tasks. A further interesting point in this connection is that the essays written by the Northeastern 13-year-olds who wrote these five difficult SC exercises ranked in either first or second place nationally among the four regions (cf. NAEP Report 10-W-02, Writing Achievement, 1969-79: Results From the Third National Writing Assessment).

---

[1] Statistically significant at the .05 level.

TABLE 3-1. Significant Differences in Correctness Percentages of Subgroups Compared With National Correctness Percentages

Exercise Title	Age	National % Correct	Region				Sex		Race		
			Southeast	West	Central	Northeast	Male	Female	White	Black	Oth
Coat	9	38	—	—	—	—	—	—	+5	-23	—
Clown	9	40	—	—	—	—	-6	+6	+5	-26	-
Lemonade	9	34	-8	—	—	+8	—	—	+4	-16	—
	13	77	—	—	—	—	-2	+2	+3	-17	—
	17	91	-5	—	—	+4	-2	+1	+4	-20	-1
Cries	9	28	—	—	—	—	—	—	+4	-18	—
	13	66	—	—	—	—	-4	+4	+5	-25	-1
	17	84	—	—	—	—	-3	+3	+3	-20	—
Rope	9	16	—	—	—	—	—	—	+2	-10	—
	13	60	—	—	—	—	-4	+4	+6	-28	—
	17	80	—	—	—	—	-3	+2	+4	-22	-1
Magician	9	16	—	—	—	+5	—	—	+2	-12	—
	13	52	—	—	—	—	-5	+5	+6	-30	—
	17	76	-5	—	—	—	-3	+3	+4	-22	-1
Troops	9	20	—	—	—	—	-4	+4	+4	-17	—
	13	52	—	—	—	—	-5	+4	+5	-26	—
	17	70	-10	+5	—	—	-4	+4	+5	-30	—
Bubble	9	18	—	—	—	—	-4	+4	+3	-15	-
	13	49	—	—	—	—	-8	+7	+5	-26	—
	17	65	-6	—	—	—	-8	+6	+4	-27	—
Forest Fires	13	40	—	—	—	—	-3	+2	+4	-19	-1
	17	62	-8	+5	—	—	-5	+4	+4	-30	—

TABLE 3-1 Continued.

Exercise Title	Age	National % Correct	Region				Sex		Race		
			South-east	West	Central	North-east	Male	Female	White	Black	Other
Guard	9	6	---	---	---	---	---	---	+1	-5	---
	13	32	---	---	---	+5	-3	+3	+4	-22	---
	17	61	---	---	---	---	-4	+4	+5	-22	---
Pebbles	9	3	---	---	---	---	---	---	+1	-3	---
	13	21	---	---	---	+4	---	---	+2	-12	---
	17	48	-6	---	+5	---	-5	+4	+4	-25	-14
Hikers	13	23	---	---	---	+5	-5	+5	+2	-11	---
	17	44	-5	---	---	---	-6	+5	+4	-23	---
Lookout	13	20	---	---	---	+4	-2	+2	+3	-14	---
	17	44	---	---	---	---	-3	+3	+4	-23	---
Bus	13	12	---	---	---	+3	-3	+3	+2	-9	---
	17	33	---	---	---	---	-6	+5	+4	-23	-10

Exercise Title	Age	National % Correct	Parental Education			Type of Community		
			Not Graduated High School	Graduated High School	Post High School	Disadvantaged Urban	Rural	Advantaged Urban
Coat	9	38	---	---	+6	-22	---	+12
Clown	9	40	-16	---	+11	-27	-11	+20
Lemonade	9	34	-18	---	+14	-22	---	+14
	13	77	-11	---	+6	-25	-13	+6
	17	91	-11	---	+5	-20	---	---

TABLE 3-1 Continued.

Exercise Title	Age	National % Correct	Parental Education			Type of Community		
			Not Graduated High School	Graduated High School	Post High School	Disadvantaged Urban	Rural	Advantaged Urban
Cries	9	28	-10	---	+9	-15	---	---
	13	66	-13	---	+9	-29	-11	+10
	17	84	-11	-4	+7	---	---	+6
Rope	9	16	-7	---	+6	-10	---	---
	13	60	-18	---	+13	-31	-11	+11
	17	80	-8	-4	+6	-18	---	+8
Magician	9	16	---	---	+5	-12	-5	---
	13	52	-18	-3	+12	-22	-10	+12
	17	76	-12	-4	+7	-17	---	+11
Troops	9	20	-14	---	+9	-16	-9	+1[illegible]
	13	52	-18	-5	+14	-23	---	+15
	17	70	-18	-8	+12	---	---	---
Bubble	9	18	-11	---	+8	-14	-6	+12
	13	49	-16	---	+10	-22	---	+14
	17	65	-17	-6	+10	-18	---	---
Forest Fires	13	40	-14	-4	+11	-16	---	+15
	17	62	-15	-6	+10	-15	-10	+13
Guard	9	6	-4	---	---	-6	-3	+6
	13	32	-16	---	+10	-23	-9	+18
	17	61	-20	-7	+11	-21	---	+9
Pebbles	9	3	-3	-2	+3	-3	---	---
	13	21	-12	---	+8	-12	-8	+8
	17	48	-25	---	+8	-27	-16	+10

TABLE 3-1 Continued.

Exercise Title	Age	National % Correct	Parental Education			Type of Community		
			Not Graduated High School	Graduated High School	Post High School	Disadvantaged Urban	Rural	Advantaged Urban
Hikers	13	23	-10	-3	+7	-17	-10	+8
	17	44	-17	-3	+7	-22	—	+10
Lookout	13	20	-12	-3	+8	-13	—	—
	17	44	-21	-5	+10	-15	—	+9
Bus	13	12	-6	—	+5	-5	-4	+7
	17	33	-12	-3	+5	-13	—	+15

Relationship to Essay Scores

With certain variations, the basic plan of the 1978-79 writing assessment as regards essays and SC tasks was that each student wrote an essay and worked three SC exercises, at least one of which was difficult. Students achieving primary-trait (PT) scores of 3 or 4 can be said to have good PT scores; those with PT scores of 1 or unscorable can be said to have poor PT scores. (See Appendix B for description of PT scoring.) Similarly, students writing acceptable combinations on all three SC exercises can be said to possess good SC skills, while those writing no acceptable combinations have poor SC skills. The question then arises, what is the relationship between good and poor SC skill, respectively, and good and poor PT scores?

Tables 3-2 and 3-3 provide data answering this question. Each table presents as a base of comparison the percentage of writers within the entire national sample who earned PT scores of 4 or 3 (Table 3-2), or PT scores of 1 or unscorable (Table 3-3), then looks within the subgroups of writers having good SC skill and poor SC skill, asking what percentages of these subgroups earned good and poor PT scores. The answers are shown as difference percentages subtracted from the national PT-score percentages. They may be understood as the percentage of writers with good or poor SC skills that surpassed or fell short of the national level of success for the PT ratings -- in Table 3-2, high PT scores; in Table 3-3, poor PT scores.

Obviously, both tables show consistently unidirectional relationships between SC ability and the ability to score well on PT writing exercises. Averaging the 12 comparisons reported in each table, one sees in Table 3-2 that people who consistently combine sentences correctly have a 12% greater tendency to achieve high PT scores, while those who cannot combine sentences at all are 12% less likely to achieve high PT scores. Conversely, Table 3-3 shows that on average, good sentence-combiners are about 16% less likely to register poor PT scores, while noncombiners are about 15% more likely to perform incompetently in their essay writing. In short, the ability to combine sentences and the ability to write essays meriting high PT scores are highly associated.

TABLE 3-2. Good/Excellent Primary-Trait Essay Scores of Writers With Good and Poor SC Skills Compared With National Good/Excellent Primary-Trait Scores

Age	Essay Title	National %, PT Score of 3 or 4	Differences From National %s of Writers With Good and Poor SC Skills	
			Writers With Good SC Skill (All SC Exercises Correct)	Writers With Poor SC Skill (No SC Exercises Correct)
9	Goldfish	13.7	+13.4	-3.5
9	Fireflies	10.0	+8.8	-1.6
9	Puppy Letter	15.6	+17.2	-7.0
9	Poster Calendar	48.9	+31.6	-14.4
13	Loss	19.8	+9.8	-8.1
13	Rainy Day	6.2	+7.1	-4.1
13	Principal Letter	20.3	+14.0	-13.0
13	Poster Calendar	79.9	+11.9	-17.3
17	Grape Peeler	15.6	+5.1	-10.0
17	Stork	74.8	+11.9	-26.6
17	Recreation Center	15.2	+6.8	-13.7
17	Electric Blanket	46.7	+9.6	-21.6
		Across-age averages =	+12.3	-11.7

TABLE 3-5. Poor Primary-Trait Essay Scores of Writers With Good and Poor SC Skills Compared With National Poor Primary-Trait Scores

Age	Essay Title	National %, PT Score of 1 Plus Unscorable Responses	Differences From National %s of Writers With Good and Poor SC Skills	
			Writers With Good SC Skill (All SC Exercises Correct)	Writers With Poor SC Skill (No SC Exercises Correct)
9	Goldfish	59.9	-27.6	+6.2
9	Fireflies	32.6	-21.6	+6.2
9	Puppy Letter	38.4	-23.9	+12.3
9	Poster Calendar	47.6	-30.6	+14.3
13	Loss	41.5	-14.5	+11.7
13	Rainy Day	33.7	-9.3	+10.9
13	Principal Letter	35.9	-18.7	+23.6
13	Poster Calendar	15.0	-10.5	+15.5
17	Grape Peeler	63.8	-7.8	+11.4
17	Stork	2.0	-0.8	+7.7
17	Recreation Center	27.3	-11.2	+28.1
17	Electric Blanket	34.8	-12.1	+26.9
	Across-age averages =		-15.7	+14.6

In addition to PT scoring, one essay at each age level was scored holistically on a four-point scale, scores of 3 and 4 being high. (See Appendix B for a description.) Holistic scoring is governed not only by the extent to which the content and form of an essay constitute appropriate responses to the given essay topic, but also by the general quality of sentence structure, vocabulary, style and mechanics present in the writing. The high-holistic scores of writers who are good and poor sentence-combiners, respectively, compared with the high-holistic scores in the entire national sample are:

Age	Essay Title	National Percentage of High-Holistic Scores (3 or 4)	Difference Scores of High-Holistic Writers With Good and Poor SC Ability	
			Good SC Ability	Poor SC Ability
9	Kangaroo	49	+36	-15
13	Describe	52	+26	-22
17	Describe	40	+15	-28
		Across-age averages =	+26	-22

Here one sees an even higher degree of agreement between SC ability and writing ability than was observed in the PT-scored essays.

Clearly, those who write good essays also tend to do well on SC tasks. Yet one cannot conclude from this fact that either ability would act as a causal factor in the development of the other. Both may be related to some third variable, such as intelligence, motivation, attention, perseverance, decenteredness, and so forth. Still, since acceptable sentence-combining can be directly taught and practiced in a way that good writing cannot be directly taught, the trend over recent years in American schools toward more widespread use of SC as one kind of regular language practice would seem, in light of these assessment findings, to deserve continuation and perhaps reinforcement.

But a further question arises: Which students at which ages would likely profit most from in-school SC practice? Looking again at Table 3-2 and the data on holistic scores, one sees that in both instances the positive differences between the good sentence-combiners' high essay scores and those of the nation decrease with age, while the negative differences of the poor sentence-combiners increase. What the high positive difference scores of the 9-year-olds reflect, since one finds very little school-sponsored SC practice before grade four (roughly age 9), is not prior SC practice but precocious natural development. These are 9-year-olds on a fast verbal-development schedule, 9-year-olds

who, compared with their agemates, are more competent sentence-combiners and better writers, and who got that way naturally, that is, without benefit of specific academic instruction. Developmentally, they compare favorably with average youngsters several years older.

On the other hand, among 17-year-olds there is the least positive difference between good sentence-combiners and good writers, and the greatest negative difference, as is indicated in both Table 3-2 and the data on holistically scored essays. Thus it seems a straightforward conclusion that the place where SC practi would produce the greatest effect upon writing ability is between the ages of 13 and 17, targeted at those students who, at age 13 or 14, show themselves on an appropriate diagnostic test (perhaps not unlike the NAEP items) to be poor sentence-combiners. Not only would this instructional policy reduce the percentage of 17-year-olds who remain poor sentence-combiners, it would also, if there is a causal relationship between SC ability and writing ability, increase the number of 17-year-olds producing high-quality essays. All of this, of course, in no way reduces the possibility that pedagogical SC may be valuable for all students at every age above grade four -- a possibility favorably indicated by the relationships discussed above, taken in toto, of SC ability to essay-writing ability.

Relationship to Essay Syntax

A third group of variables to which SC ability is assumedly related consists of those characterizing the grammatical structure of the sentences in student essays. Indeed, research on the effectiveness of pedagogical SC invariably uses counts of the syntactic structures comprising free student writing as its dependent variable. The parameters of normal syntactic development in the writing of school-age children have been widely studied over the past 20 years, following the advent of transformational theories of sentence structure, and SC research typically evaluates this or that approach to pedagogical SC in terms of the direction and extent to which it acts as a biasing factor enhancing normal development of syntactic fluency in any of a large number of possible ways, in various modes of discourse. The question here is: What relationship exists between SC ability as measured by the National Assessment exercises and level of syntactic maturity as measured by structure counts in the NAEP essay exercises?

Segmentation of writing into T-units, followed by an exhaustive syntactic analysis of each, is quite time-consuming and costly. Interested primarily in comparisons across time, and able thusfar to analyze only one exercise at each age level, National Assessment chose for its first venture into explicit

syntactic analysis one expressive essay at each age that had been administered in both of the earlier writing assessments (1978-79 and 1973-74) and was scheduled for use again in the 1978-79 assessment. These were "Fireflies" at age 9, "Rainy Day" at age 13 and "Stork" at age 17. Readers are referred to National Assessment Reports 10-W-01, 10-W-02 and 10-W-03 for a description of these exercises and for tabular summaries of the resultant T-unit and T-unit-constituent counts. (These reports, the first ever to yield detailed syntactic analyses of a national sample cross-sectioned at three points over a 10-year period, reveal uncanny numerical identities in the structure counts from year to year at all three ages, betokening the apparently bedrock immutability over time of preferred syntactic structure when age and topic are held constant.)

Unfortunately, although there is much room for further examination in the vast body of data accumulated in National Assessment's syntax analysis, the expressive exercises just mentioned do not yield the kind of writing in which one would ideally look for relationships between mature SC ability and sentence syntax. The "Stork" exercise at age 17, for instance, typically took the form of a playful narrative anecdote featuring short sentences of exclamatory dialogue. Nearly all the writers selected a stylistic register that precluded the kind of syntactic elaboration one might see in narrative sentences conveying sequences of chained events, reports of characters' mental actions or detailed descriptions of settings and events -- in short, a stylistic register precluding exactly the kinds of sentences the writers of "Stork" were asked to produce in the two difficult SC exercises they wrote, "Pebbles" and "Hikers."

What can be reported, however, are the findings of an examination of SC ability and written syntax at age 17 in the third writing assessment conducted not by National Assessment itself, but by a team of researchers funded by the National Institute of Education, with NAEP's approval and cooperation (Diehl, 1980). In this study, the researchers selected a subsample of 160 essays written by 17-year-olds in response to the persuasive argumentation exercise "Recreation Center" (described in the National Assessment Report 10-W-03). This subsample consisted of four sets of 40 essays each, drawn randomly from those in the NAEP national sample scoring at each of the four PT score points -- 40 papers scoring 4, 40 scoring 3, and so on. The writing was segmented into T-units and analyzed syntactically. Average T-unit length did not differ significantly across the four score points, a result consonant with repeated findings of developmental research that writers' expository T-units stabilize in length from grade 12 on. The pattern of embeddings and the average clause length in the four sets of papers were as follows:

Essays Grouped by PT Score	Embeddings Per 100 T-Units			Average Clause Length, Words Per Clause
	Clausal	Phrasal	Total	
4 (good)	65	65	130	9.7
3	85	65	150	8.7
2	85	63	148	8.4
1 (poor)	95	55	150	8.0

Here the groupings of essays by PT score indicates greater syntactic maturity associated with the higher PT scores. Developmental research shows that from age 17 or 18 on, clause length (rather than T-unit length) is the main indicator of maturity, in that the number of unreduced (finite-verb) clauses per T-unit decreases as writers learn to reduce more of their clauses to phrase forms. The above data are consistent with these developmental facts. While the 1, 2 and 3 papers contain almost exactly the same number of embeddings per T-unit, the shift away from clauses to phrases is obvious in the better papers, as are the longer clause lengths yielded thereby. The 4 papers have fewer clause embeddings still, and fewer total embeddings. Their longer clause length necessarily results from longer phrases and greater use of coordination (as opposed to embedding) as a sentence-combining strategy -- both hallmarks of maturity.

So much for background. Here the study poses the question from which its interest derives: What happens to embedding patterns and clause lengths when the essays are grouped not by PT score but by their writers' performance on the National Assessment SC exercises? Included among the three SC tasks required of the students who wrote "Recreation Center" was the difficult exercise "Forest Fires." Two groups of "Forest Fires" responses were identified: "good quality," defined as a subset of all acceptable responses, and "poor quality," defined as a subset of unacceptable responses. Specifically, good-quality responses were defined as those retaining the explicit causal statement given in the exercise stem, "careless people often cause forest fires." The three response patterns fitting this specification observed in the 160-student sample were:

1. Two relative clauses:
   Careless people who tend to drop lighted cigarettes often cause forest fires that/which can destroy lives and property. (22 instances)

2. Main clause passivized:
   Forest fires that/which can destroy lives and property are often caused by careless people who tend to drop lighted cigarettes. (8 instances)

3. Participial conjoining with relative clause:
   Careless people tend to drop lighted cigarettes causing forest fires that/which can destroy lives and property. (6 instances)

Poor-quality responses, in turn, were those in the following categories:

1. Garbled:
   This category included sentences with all lexical items present, but with syntactic alterations and ineptitudes. It also included sentences scored acceptable that were stylistically weak, such as "People who are careless often destroy lives and property by dropping lighted cigarettes to cause forest fires." (24 instances)

2. More than one sentence:
   This category included lexically and syntactically acceptable combinations of more than one T-unit, which if correctly punctuated would be more than one sentence. (4 instances)

3. Compound sentences with repeating nouns requiring subordination:
   This category included combinations of more than one T-unit, such as "Careless people often cause forest fires, tend to drop lighted cigarettes and forest fires can destroy lives and property." (14 instances)

Altogether, 78 of the 160 students were identified as poor or good sentence-combiners by these criteria. Their essays were regrouped accordingly, and their syntax counts recomputed, with the following rather surprising results:

	Essays by PT Score				Clause Embeddings Per 100 T-Units				Average Clause Length, Words
	1	2	3	4	Noun	Adjective	Adverb	Total	
Good-quality SC: (n=36)	2	4	10	20	27	18	19	64	10.2
Unacceptable SC: (n=42)	22	12	6	2	25	14	38	77	8.7

One first notes that the regrouping by SC ability finds essays from all four PT score points in each category, while the skewed distribution within each reflects the SC-PT relationship discussed in the preceding section of this report. Second, one sees that the reduction in total clauses results from a dramatic decline in number of adverbials, coupled with slight increases in nominals and ajectivals. In developmental research generally, this is exactly the pattern by which total clausal embeddings decrease somewhat, then stabilize in expository writing beyond grade 12. Third and most notably, the average clause length of the good sentence-combiners is 10.2, considerably more mature than the 9.7 average of all PT-score 4 writers. Developmental norms backgrounding these figures show that the clause length of expository writing typically increases from 8.5 words in grade 12 to 11.5 words among skilled adults. What is indicated is that good SC ability as defined here contributes a factor to maturity of written syntax that is not measured by PT score. If confirmed by subsequent more broadly based research, these results constitute potentially strong evidence supporting, one, the curricular validity of pedagogical SC of the kind that constrains writers to produce combinations preselected for high maturity level and good quality; and two, the likelihood that good quality responses to SC exercises used on a test or assessment basis will yield an efficient proxy measure of attained syntactic maturity.

CHAPTER 4

SUMMARY DISCUSSION

This report began with the observation that NAEP's 1978-79 assessment of SC skills represents the first use anywhere of sentence combining as a task-type in a widescale writing assessment. In light of the results presented in this report, it now seems fair to conclude that SC tasks prove quite effective in this regard, and exhibit a potential for even more refined and highly specified uses in the measurement of writing skill.

The facts reported here indicate that SC exercises used for assessment can be successfully introduced to students who assumedly lack prior experience with pedagogical SC, if the directions used refer to the activity as if it were an editing task, and if no artificial cuing apparatus is included. A two-sentence combination plus its solution used as an illustration, followed by one or two easy exercises in the form of three-sentence combinations, will familiarize students with the SC process enough to enable them subsequently to attempt whatever difficult tasks the assessor wishes to give them.

The grammatical scope of National Assessment's first battery of SC exercises remained of necessity somewhat limited, since the format used applied mainly to the syntax of noun modification, that is, relative clauses and their prenoun and postnoun word and phrase reductions. But this is not to say that other schemes may not be developed in the near future, including direction sequences that phase in artificial cues on a stepwise basis as students work their way through a series of differently formated exercises. There is a need here for some small-scale feasibility research probing the limits of complexity that may be reached before exercise directions lose the self-evident quality required for assessment purposes. The directions and exercise format used by National Assessment represent a beginning only, and in all likelihood will rather quickly be extended and improved upon by clever researchers.

Another question that measurement research will shortly answer concerns the extent to which, if at all, prior acquaintance with pedagogical SC affects a student's performance when working SC exercises on tests or assessments. Except for the roughly 30% of the 9-year-olds who are still too cognitively immature to produce scorable responses to even the easiest items, nearly all

students could solve easy SC problems introduced by the NAEP-designed directions. But the difficult exercises proved quite another matter. As suggested earlier in this report, while easy (three-sentence) items may seem like straightforward editing tasks to all students whether or not they are familiar with pedagogical SC, difficult items (five sentences or longer) are so unlike any editing activities persons encounter in or out of school that no student will think of them in that light. Yet students who have had experience with pedagogical SC will recognize at once what is expected of them, whereas students who have not may become confused and daunted. If and when SC becomes more or less a standard language-learning activity in schools, the effect of difficult SC assessment exercises will be felt the same by all students, and results will vary only in accordance with their syntactic ability. Until then, it would seem desirable to include in any SC assessment an item or two eliciting background information on each student's prior experience with SC that could be factored into the exercise results. In the meantime, researchers will wish to characterize in empirical terms what effect presence or absence of prior SC experience exerts upon the working of SC exercises used in assessments.

Another important aspect of the SC data reported here is the clarity with which three separate sources of difficulty in any SC task reveal themselves. These are:

1. Number of sentences (hence, amount of information) to be combined.

2. Number of possible syntactic relationships into which the sentences may enter one with another.

3. Degree of likelihood that students at any age control the syntactic transformations necessary to make the combination.

To reiterate, "Bubble," "Guard" and "Pebbles" illustrate exercises whose difficulty arises primarily from number of sentences to be combined; the relationships are straightforward and the transformations well controlled. "Bus" illustrates an exercise consisting of a set of sentences that permit a great many possible interrelationships. "Forest Fires" contains only three sentences and their relationships are obvious, but by inviting two unreduced relative clauses calls for transformations that are difficult for young writers to perform. (The difficulty of "Hikers" and "Lookout" arises from rather evenly distributed combinations of causes.) This three-part analysis of the source of difficulty in SC exercises, while it contains nothing that psycholinguists studying language production have not examined in considerably greater detail, would represent a step ahead if applied to pedagogical SC research. Presumably, too, it will be

built upon and refined in subsequent uses of SC items in writing assessments, and may lead to more highly discriminant measures of the syntactic fluency of individual student writers.

Still other researchers may wish to pursue the speculation voiced earlier in this report that specially selected SC tasks might serve to identify latent talent, or at least a penchant, for fine or artistic writing. Examples from the battery of exercises reported here, as noted earlier, would be the adjectives in "Cries" cast as coordinate appositives in either the introductory or postnoun position, or the double nominative-absolute phrases in "Bus." Researchers might experiment with exercise directions containing illustration sentences combined artistically, as a way of signaling the fact that a fine-writing response is wanted without having to attempt the impossible task of defining what is meant by the notion of fine or artistic writing. One would also wish to examine the prose style of writers who consistently produce fine-writing responses to SC exercises, relating their SC performance to the structures they select in their free writing.

Even more important will be to look more closely at the relationship between acceptable performance on SC tasks and level of syntactic fluency in writing, by following up on schemes that subdivide acceptable SC responses into categories. A three-part scheme might be tried, such as:

1. High quality (including but not limited to fine-writing patterns).

2. Acceptable but undistinguished.

3. Stylistically weak.

The discovery of positive correlations between quality-level of SC response and observed syntactic-fluency level would give strong support to the idea, currently out of favor among a majority of persons involved with pedagogical SC, that single-sentence exercises cued to yield mature syntactic patterns are every bit as important as are uncued whole-discourse (multisentence) exercises. On the developmental side, it may be found that an approximately constructed battery of SC exercises might prove workable as a valid proxy measure of attained syntactic fluency, more or less along the lines pioneered by Kellogg Hunt and his associates with their whole-discourse rewriting exercises. An empirically selected mix of individually validated single-sentence SC items may well constitute a considerably more reliable measure of syntactic fluency than any whole-discourse exercise that might be found.

Still other ideas for related research will occur to readers of this report. It is not too much to imagine, for example, that

persons might develop diagnostic tests of SC skill, perhaps sensitive to the three sources of exercise difficulty discussed above (information load, semantic relationships and transformations), to be administered in grades 9 or 10, whose results would indicate which students would profit most from what particular forms of pedagogical SC practice during their remaining high school years.

Summing up, the results of the battery of SC exercises included in National Assessment's 1978-79 writing assessment cast further light on sentence combining as a separately identifiable skill constituent in, although seldom overtly recognized as a part of, what is termed general writing ability. Performance on SC exercises varies with age and membership in the various demographic subgroups, in ways resembling performance on more familiar kinds of academic exercises. SC exercises can be used successfully in assessment settings whether or not students have had prior experience with SC within their school programs, although it remains to be determined whether and to what extent familiarity with pedagogical SC may affect performance on a SC assessment. The 1978-79 results provide baseline data for subsequent assessments of SC skill. The ability to combine sentences is associated with the overall quality of student writing as measured by both holistic and primary-trait scoring. Proposals abound for further research on SC in an assessment setting, including variations of exercise format and the range of grammatical structures assessed, and the use of SC to identify potential fine-writing talent and to provide a proxy measure of syntactic fluency in free writing. Thus the usefulness of SC as an assessment item-type seems well established, and its future promising.

APPENDIX A

FREQUENCY RANKINGS OF CORRECT PAPERS

TABLE A-1. Frequency Ranking of Patterns of Correct Responses, With Percentage of Occurrence Among Total Correct Responses (Reported Instances = >10)

Exercise	Age 9 Rank	Age 9 %*	Age 13 Rank	Age 13 %*	Age 17 Rank	Age 17 %*	Response Pattern
Coat	1	95	-	-	-	-	Bill's new leather coat was in the closet.
	2	2	-	-	-	-	Bill's new coat was leather and was in the closet.
	3	1	-	-	-	-	Bill's coat was a new one and leather and was in the closet.
Clown	1	88	-	-	-	-	The jolly clown was smoking a fat cigar.
	2	3	-	-	-	-	The clown was jolly and was smoking a fat cigar.
	3	2	-	-	-	-	The clown was smoking a fat cigar and was jolly.
Lemonade	1	55	1	76	1	82	The barefoot boys drank the cold lemonade.
	6	4	3	4	2	6	The boys who were barefoot drank the cold lemonade.
	2	17	2	6	3	3	The barefoot boys drank the lemonade cold.
	5	4	6	1	4	1	While barefoot, the boys drank the cold lemonade.
	3	7	4	3	6	1	The boys were barefoot and drank the cold lemonade.
	4	5	5	2	5	1	The barefoot boys drank the lemonade while it was cold.
Shoes	1	52	1	70	1	87	Her thin small shoes were lost in the storm.
	2	39	2	15	2	7	Her shoes were thin and small and were lost in the storm.
	-	-	4	1	3	2	Her shoes, thin and small, were lost in the storm.
	3	2	3	1	4	1	Her shoes were so thin and small that they were lost in the storm.
	-	-	5	1	5	1	Because they were thin and small, her shoes were lost in the storm.
Rope	1	63	1	69	1	73	A twisted rope hanging from a tree branch was a clue to the mystery.
	2	13	2	20	2	20	A twisted rope that was hanging from a tree branch was a clue to the mystery.
	-	-	5	1	3	2	A clue to the mystery was a twisted rope that was hanging from a tree branch.
	3	7	3	2	4	2	The rope that was a clue to the mystery was twisted and hanging from a tree branch.
	4	5	4	2	6	1	A twisted rope was hanging from a tree branch and was a clue to the mystery.
Magician	1	76	1	85	1	93	John knows a clever magician who can make an elephant disappear.
	2	20	2	10	2	4	John knows a magician who is clever and can make an elephant disappear.
	-	-	4	1	3	1	John knows a magician who is clever because he can make an elephant disappear.
	-	-	3	1	4	0	John knows a magician so clever that he can make an elephant disappear.
Troops	1	87	1	90	1	88	The strong fearless captain took care of his tired hungry troops.
	2	3	2	2	2	3	The strong fearless captain took care of his troops, who were tired and hungry.
	3	3	3	2	3	2	The strong fearless captain took care of his troops, because/when they were tired and hungry.
	-	-	4	1	4	1	The captain, who was strong and fearless, took care of his tired hungry troops.
	-	-	4	1	6	1	The captain was strong and fearless and took care of his tired hungry troops.
	-	-	-	-	5	1	The captain, being strong and fearless, took care of his tired and hungry troops.
Bubble	1	76	1	72	1	76	The plants are kept dry by a large plastic bubble that covers the entire garden.
	3	5	1	7	2	10	A large plastic bubble that covers the entire garden keeps the plants dry.
	2	8	3	6	3	5	A large plastic bubble covers the entire garden and keeps the plants dry.
	4	4	4	2	4	4	A large plastic bubble covers the entire garden to keep the plants dry.
	-	-	5	1	5	2	The plants are kept dry by a large plastic bubble covering the entire garden.
	-	-	-	-	6	1	A large plastic bubble covering the entire garden keeps the plants dry.
	-	-	-	-	7	1	The bubble that keeps the plants dry is large and plastic and covers the entire garden.
Forest Fires	-	-	1	43	1	46	Careless people tend to drop lighted cigarettes which often cause forest fires that can destroy lives and property. / Careless people who tend to drop lighted cigarettes often cause forest fires that can destroy lives and property.
	-	-	2	21	2	12	Careless people tend to drop lighted cigarettes and often cause forest fires that can destroy lives and property.
	-	-	3	16	3	10	Careless people tend to drop lighted cigarettes which often cause forest fires and destroy lives and property.
	-	-	5	9	4	6	Careless people often cause forest fires by dropping lighted cigarettes, which can destroy lives and property.
	-	-	8	1	5	5	Forest fires that can destroy lives and property are caused by careless people who tend to drop lighted cigarettes.
	-	-	-	-	6	4	Dropping lighted cigarettes, careless people often cause forest fires that can destroy lives and property.
	-	-	6	5	7	3	People often cause forest fires that can destroy lives and property by carelessly dropping lighted cigarettes.
	-	-	4	10	8	2	Careless people tend to drop lighted cigarettes, often cause forest fires, and destroy lives and property.
	-	-	7	3	9	1	By dropping lighted cigarettes and causing forest fires, careless people can destroy lives and property.
	-	-	9	1	-	-	When careless people drop lighted cigarettes and cause forest fires, they can destroy lives and property.
	-	-	9	1	-	-	Careless people, in/by dropping lighted cigarettes, often cause forest fires that can destroy lives and property.

*These are percentages of the percentages of correct responses. See Table 2-1 for percentages of correct responses upon which these figures are based.

TABLE A-1 Continued.

Exercise	Age 9 Rank	Age 9 II	Age 11 Rank	Age 11 II	Age 13 Rank	Age 13 II	Sentence Pattern
Guard	1	33	1	64	1	61	A bored guard at the doorway kept the children from touching the dusty stuffed animals in the museum display.
	-	-	6	3	2	13	A bored guard at the doorway kept the children from touching the dusty stuffed animals that were in the museum display.
	-	-	4	3	3	4	A bored guard that was at the doorway kept the children from touching the dusty stuffed animals that were in the museum display.
	-	-	3	4	4	3	A bored guard that was at the doorway kept the children from touching the dusty stuffed animals in the museum display.
	-	-	-	-	5	2	A bored guard was at the doorway, keeping the children from touching the dusty stuffed animals in the museum display.
	-	-	-	-	6	1	A bored guard, by being at the doorway, kept the children from touching the dusty stuffed animals in the museum display.
	2	13	2	7	-	-	A bored guard at the doorway kept the children from touching the dusty stuffed museum display animals.
Pebbles	1	15	1	24	1	26	The shiny yellow pebbles gleaming like cats' eyes marked the path to a magic underground kingdom ruled by a wizard.
	1	16	4	11	2	17	The shiny yellow pebbles that gleamed like cats' eyes marked the path to a magic underground kingdom ruled by a wizard.
	-	-	2	16	3	13	The shiny yellow pebbles that were gleaming like cats' eyes marked the path to a magic underground kingdom that was ruled by a wizard.
	-	-	6	6	4	10	The shiny yellow pebbles gleaming like cats' eyes marked the path to a magic underground kingdom that was ruled by a wizard.
	-	-	3	9	5	10	The shiny yellow pebbles that gleamed like cats' eyes marked the path to a magic underground kingdom ruled by a wizard.
	-	-	3	13	6	8	The shiny yellow pebbles that were gleaming like cats'eyes marked the path to a magic underground kingdom ruled by a wizard.
Hikers	-	-	1	81	1	84	The hikers tramped along the steep narrow rocky path that curved upward toward the mountain that appeared ahead through the clouds.
	-	-	4	3	3	3	The hikers tramped along the steep narrow rocky path curving upward toward the mountain that appeared ahead through the clouds.
	-	-	-	-	3	3	Because it was steep, the hikers tramped along the narrow rocky upward-curving path toward the mountain that appeared ahead through the clouds.
	-	-	-	-	4	2	The hikers tramped along the steep narrow rocky path as it curved upward toward the mountain that appeared ahead through the clouds.
	-	-	4	2	5	2	The hikers tramped along the steep narrow rocky path that curved upward toward the mountain as it appeared ahead through the clouds.
	-	-	-	-	6	1	The hikers tramped along the narrow rocky path that curved steeply upward towards the mountain that appeared ahead through the clouds.
	-	-	2	3	-	-	The hikers tramped along the steep narrow rocky path that was curving upward toward the mountain that appeared ahead through the clouds.
	-	-	3	3	-	-	The path that the hikers tramped along was steep and curved upward toward the mountain that appeared ahead through the clouds.
Lookout	-	-	2	23	1	24	The frightened lookout clinging to the mast realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	1	25	2	21	The frightened lookout clung to the mast as he realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	3	2	3	8	The frightened lookout who was clinging to the mast realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	7	1	4	3	The frightened lookout clinging to the mast realized the tidal wave would swamp the ship, sending it plunging to the depths.
	-	-	7	1	5	6	The lookout clinging to the mast was frightened as he realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	-	-	6	4	Clinging to the mast, the frightened lookout realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	8	4	7	4	The lookout was frightened and clung to the mast as he realized the tidal wave would swamp the ship and send it plunging to the depths.
	-	-	4	8	8	4	The frightened lookout clung to the mast, realizing the tidal wave would swamp the ship and send it plunging to the depths.
			-	-	9	3	Frightened as he clung to the mast, the lookout realized the tidal wave would swamp the ship and send it plunging to the depths.
			-	-	10	2	The frightened lookout clung to the mast as he realized the tidal wave would swamp the ship, sending it plunging to the depths.
			5	4	-	-	The frightened lookout clung to the mast and realized the tidal wave would swamp the ship and send it plunging to the depths.
Bus	-	-	1	14	1	19	The people standing on the corner with cold hands dangling at their sides were waiting for a bus that was already 20 minutes late.
			2	6	2	24	With their cold hands dangling by their sides the people stood on the corner, waiting for a bus that was already 20 minutes late.
			2	9	3	10	The people with cold hands dangling by their sides were standing on the corner waiting for a bus that was already 20 minutes late.
			3	3	4	3	There were people who were standing on the corner dangling their cold hands by their sides while waiting for a bus that was already 20 minutes late.
			-	-	3	3	The people, whose cold hands dangled by their sides, stood on the corner waiting for a bus which was already 20 minutes late.
			-	-	6	2	The people standing on the corner waiting for a bus that was already 20 minutes late had cold hands dangling by their sides.
			5	3	7	2	Waiting for a bus that was already 20 minutes late, the people stood on the corner with their cold hands dangling by their sides.
			-	-	8	2	The people stood on the corner, their cold hands dangling by their sides, the bus they were waiting for already 20 minutes late.
			-	-	9	2	The people, whose hands were cold and dangling by their sides, stood on the corner waiting for a bus that was already 20 minutes late.
					10	1	The people stood on the corner with their cold hands dangling by their sides, waiting for a bus that was already 20 minutes late.
			4	4	-	-	The people standing on the corner had their cold hands dangling by their sides, waiting for a bus that was 20 minutes late.

*These are percentages of the percentages of correct responses. See Table A-1 for percentages of correct responses upon which these figures are based.

TABLE A-2. Percentages of Students Writing Most Frequent Combination

	Age 9	Age 13	Age 17
Coat	33%	—	—
Clown	33	—	—
Lemonade	17	59%	74%
Cries	13	50	72
Rope	9	47	58
Magician	11	59	70
Troops	16	45	60
Bubble	13	35	47
Forest Fires	—	14	29
Guard	3	20	37
Pebbles	3	5	12
Hikers	—	19	36
Lookout	—	4	10
Bus	—	2	6

APPENDIX B

THE WRITING EXERCISES, SCORING APPROACHES AND DESCRIPTIVE ANALYSES

Details of National Assessment exercise development procedures appear in the NAEP Report 10-W-40, Procedural Handbook: 1978-79 Writing Assessment (1980). Complete documentation of all exercises released after the third assessment of writing, including scoring guides and sample responses, is contained in The Third Assessment of Writing: 1978-79 Released Exercise Set (1980).

The writing exercises mentioned in Chapter 3 of this report were created by experienced writing educators. Then they were field tested, refined and reviewed carefully before being used. Each assessment contained exercises assessing several kinds of discourse on the grounds that students may be proficient in some kinds of writing but not in others.

Several types of scoring and analysis went into the creation of the essay data given in this report. Each is briefly described below. Readers desiring more information about these procedures should consult the handbook and exercise set cited above, as well as Mullis (1980), Mullis and Mellon (1980), Brown (1979), which also cite additional references. For each procedure, raters scored a random mixture of papers collected from the different assessments. Each kind of scoring was done by a different group of scorers.

Holistic Scoring

When readers holistically score papers, they do not focus upon particular aspects of a paper such as mechanics or ideas or organization. Rather, they concentrate upon forming an overall impression of each paper relative to the other papers they have read. Their primary task is to rank order the papers from best to worst, not to identify errors or to specify writing problems.

Holistic scoring involved several steps. First, the table leaders -- all of whom were experienced holistic readers -- surveyed the pool of papers from all three assessments and selected examples of papers representing four levels of quality. Then, they developed guidelines describing each level of quality and how to distinguish between top-half and bottom-half papers.

The scoring session began with some discussion of the characteristics of the anchor papers and guidelines, and then several practice scorings of other papers to refine the scoring scale description and iron out discrepancies among readers. When all readers were comfortable with the guidelines, they scored papers for an hour, after which they discussed more anchor papers. Throughout the subsequent scoring, there were periodic discussions of papers to insure that readers continued to hold to the same standards.

The Primary-Trait Scoring System (PTS)

The primary-trait approach to essay evaluation involves isolating an important writing skill, developing a task to measure it and articulating four levels of proficiency. When a reader is rating papers for PTS, he or she is rating each paper against criteria spelled out in the scoring guide instead of rating each paper in terms of the entire pool of papers. Thus, whereas a holistic scoring aims to distribute a pool of papers over a "bell shaped curve," a PTS scoring will only distribute papers according to their relationship to the scoring criteria. If none of the papers meet the criteria for the highest rating, then so be it; the object is to describe the papers, not rank order them.

Holistic scoring enables one to determine if a group of papers written at one time is better than a group written at another time but it does not provide much specific information about how the two groups differ. Primary-trait scoring provides specific information about particular rhetorical aspects of papers, but does not provide information about overall quality. Thus, it is useful to do both kinds of evaluation whenever possible.

Training for PTS scoring involves thorough discussion of the writing assignment, scoring guide and sample papers. If the assignment has been constructed to elicit evidence of proficiency in a particular writing skill, it should explicitly establish the writing situation, specifying the purpose of the communication, the audience and what must be accomplished. The instructions should unambiguously tell the writer what is required, and the scoring guide should unambiguously define four levels of proficiency in the primary skill being assessed. Generally, level "1" indicates no evidence of the skill; level "2," marginal evidence; level "3," solid performance; and level "4," very good performance. Scorers discuss each level and study papers exemplifying each until everyone feels comfortable with the system. Then scoring commences, with periodic discussion of troublesome papers. All papers were rated independently by two scorers, with disagreements being reconciled by a third.

Syntax and Mechanics

In addition to being rated for quality, the "Stork" and "Describe" papers were also analyzed in terms of their syntactic and "mechanical" features. Syntax refers to the ways in which words are put together to form phrases, clauses and sentences. Mechanics refers to the ways in which writers handle basic conventions of writing such as punctuation, spelling or word choice. A syntactic analysis involves breaking each paper up into its T-units (a T-unit is a main clause with all its attendant modifying words, phrases and dependent clauses) and examining the ways in which writers embed information in T-units and join T-units together. A mechanics analysis involves classifying the kinds of errors writers make in sentence use, punctuation, spelling, and so forth. Both kinds of analysis were done by experienced English teachers thoroughly trained in grammar, usage and linguistics.

BIBLIOGRAPHY

Brown, Rexford G. "Evaluations of Writing: Some Theoretical Considerations and Practical Suggestions." St. Louis: CEMREL, 1979.

Diehl, P. et al. Correlation of Syntactic/Stylistic Fluency and Judgment With Rhetorical Fluency and Judgment in the Persuasive Writing of 17-Year-Olds. Final Pilot Study report submitted to the National Institute of Education. Iowa City, Iowa: University of Iowa, 1980.

Mullis, Ina V.S. "Using the Primary Trait System for Evaluating Writing," no. 10-W-51. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980.

Mullis, Ina V.S. and John C. Mellon. "Guidelines for Three Ways of Evaluating Writing: Syntax, Cohesion and Mechanics," no. 10-W-50. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980.

Procedural Handbook: 1978-79 Writing Assessment, Report no. 10-W-40. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ERIC no. ED 199 263. ISBN 0-89398-397-7.

The Third Assessment of Writing: 1978-79 Released Exercise Set, no. 10-W-25. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ERIC no. ED 205 583. ISBN 0-89398-394-2.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume I -- 17-Year-Olds, Report. no. 10-W-01, 1969-70, 1973-74 and 1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ERIC no. ED 196 042. ISBN 0-89398-400-0.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume II -- 13-Year-Olds, Report no. 10-W-02, 1969-70, 1973-74 and

1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ERIC no. ED 196 043. ISBN 0-89398-401-9.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume III -- 9-Year-Olds, Report no. 10-W-03, 1969-70, 1973-74 and 1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ERIC no. ED 196 044. ISBN 0-89398-403-5.